



# VU Research Portal

## Soft econometrics as a tool for regional discrepancy analysis

Nijkamp, P.; Rietveld, P.

1981

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Nijkamp, P., & Rietveld, P. (1981). *Soft econometrics as a tool for regional discrepancy analysis*. (Serie Research Memoranda; No. 1981-19). Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

Free University  
Amsterdam  
Dept. of Economics

Peter Nijkamp  
Piet Rietveld

SOFT ECONOMETRICS  
AS A TOOL FOR  
REGIONAL DISCREPANCY ANALYSIS

Researchmemorandum 1981-19

September 1981

Paper presented at the European  
Conference of the Regional Science  
Association, Barcelona, August 1981



## Contents

Spatial Equity Problems in an Uncertain World

Soft Statistical Methods

Soft Explanatory Models

Discrepancy Analysis by means of Soft Statistics

Discrepancy Analysis by means of Soft Econometrics

A Multivariate Analysis of Dutch Interregional Welfare Discrepancies

Conclusion

## Summary

In this paper a brief overview is given of soft (qualitative, non-metric or ordinal) statistical and econometric methods in the context of regional discrepancy analysis.

Several methods are applied to Dutch interregional inequalities concerning socio-economic conditions, environmental quality and infrastructure. Special attention is paid to the extent to which the results based on an ordinal data pattern are similar to the results based on the corresponding cardinal data.



## Spatial Equity Problems in an Uncertain World

In the past, the majority of spatial economic analyses has employed a frame of reference based on the efficiency paradigm: optimal spatial allocation patterns, cost-minimizing transportation flows, maximum benefit-cost ratios, entropy-maximizing spatial flows, full employment schemes, and so forth.

Equity elements - though sometimes considered to be very important - have, in general, played a less crucial role. Several authors have pointed out the relevance of the efficiency-equity dilemma by means of the so-called transformation frontier, but in many cases equity aspects have never played a fully integrated role in spatial economic analysis. Important reasons for this neglect of distributional aspects of economic developments are:

- the major orientation of the industrialized world to the economic growth paradigm.
- the lack of an analytical framework for integrating the different dimensions implied by efficiency and equity.
- the soft (i.e., non-quantitative) nature of many equity problems (including socio-psychological perceptions, social and spatial spill-over effects, and intangible environmental impacts).
- the lack of a normative theory that may constitute a frame of reference for judging equity aspects (especially in a multi-region or multi-group context).

In conclusion, equity problems in spatial economic planning theory have often had a fuzzy nature characterized by not sharply demarcated priority sets defined on imprecise spatial impacts and based on an unsatisfactory conflict analysis. Consequently, decision-makers are often incapable of discriminating between alternative distributional states of the system at hand. This situation forms a sharp contrast with usual efficiency theories on optimal decision-making and optimal choice strategies, and it questions also the existence of a coherent and consistent choice behaviour (cf. Ponsard, 1979, and Whalen, 1980).

Following Zadeh (1973), one may make the following sequence for discrepancy analysis in a fuzzy or soft context:

- definition and use of soft or fuzzy variables in place of or in addition to cardinal variables.
- statistical characterisation of the configuration of the system concerned (for instance, via inequality measures).
- explanatory econometric analysis of relationships among variables in the system at hand.

The present paper will primarily focus on two aspects of equity problems:

- the construction of an operational framework for studying spatial discrepancies, based on a multi-dimensional profile representation of relevant variables.
- the development of a new set of statistical and/or econometric methods which are appropriate for dealing with soft variables (soft variables are measured in a non-cardinal -particularly, ordinal, nominal or categorical- metric system), so that imprecise information on spatial discrepancies can be taken into account.

A special feature of this paper will be the fact that the sensitivity of the results for the level of measurement will be analyzed. This will be done by developing a set of cardinal multivariate methods as well as their ordinal analogues, and by applying both approaches to data from the same source. The results of this discrepancy analysis will be illustrated by means of an application to Dutch regional data. Prior to the

development and application of these methods, however, a brief survey of existing techniques for soft data will be given.

### Soft Statistical Methods

Most modern statistical methods in social sciences (sociology, psychology, geography, economics etc.) are based on metric information measured on a ratio or an interval scale. The development of advanced statistical techniques, however, has not kept pace with the data base for employing such techniques. Despite the improvement of information systems, much information is either fuzzy or non-metric in nature. The existence of such soft information may inter alia lead to the problem of omitted variables or false specifications. Even many metric data may essentially be pseudo-metric because of significant uncertainties and measurement errors.

It is a common practice to introduce proxy variables for phenomena that can hardly be measured on a metric scale (for instance, happiness, satisfaction, quality of life), so that a cardinal value can be assigned to such non-metric variables, but Adelman and Morris (1974) rightly point out that this operation may easily lead to a biased view of reality.

There are several reasons why metric data are not always available in social sciences:

- the existence of significant measurement errors
- the existence of unreliable statistics
- the non-metric nature of many phenomena (scenic beauty, e.g.)
- the inability of the human mind to express priorities and perceptions on a cardinal scale.

Therefore, soft statistical methods may be useful tools to deal with inaccurate information. Such information may relate to ordinal, categorical (dichotomous or polychotomous), nominal or fuzzy information regarding phenomena to be studied.

In the past, several statistical techniques have already been developed to tackle various kinds of soft data problems, as will be shown by the following brief sample of soft statistical methods.

A traditional way of dealing with soft data (particularly, ordinal information) is the use of rank correlation analysis based on non-parametric statistics (cf. Siegel, 1956). The most well-known rank correlation coefficients are the Spearman and the Kendall rank correlation coefficient (cf. Ryans and Srinivasan, 1979). It should be noted that these correlation coefficients are only simple correlation coefficients; multiple rank correlation coefficients as an analogon to cardinal multiple correlation coefficients are hard to derive in a consistent way (see also Blalock, 1976; Hawkes, 1971; Namboodiri et al., 1975; Nijkamp and Rietveld, 1981, and Ploch, 1974). In addition, rank correlation coefficients are sometimes based on non-permissible numerical operations on ordinal data, so that one should be aware of the severe limitations and stringent assumptions of rank correlation analysis.

Path models provide also a way of treating soft information. They may be based on the assumption that the variables of direct interest cannot be observed directly, so that proxy variables reflecting a certain quantitative attribute of the original variables have to be used (see Blalock, 1964). Such path models are mainly based on correlations between qualitatively-oriented clusters of variables (for instance, as a result of principal component analyses).

A specific kind of path models is the so-called partial least squares approach (see among others Apel, 1978; Hui, 1978; and Wold, 1975, 1977, 1979). This approach aims at identifying a block structure for latent (indirectly observed) variables and their indicators (the external structure), as well as between the latent variables in each block (the internal structure) on the basis of multivariate techniques, especially iterative least squares methods. Such models may be helpful to find a certain structure in a set of latent variables, but they do not provide a solution for variables which can only be measured in an ordinal sense.

An extremely useful method for ordinal data analysis appears to be multidimensional scaling (MDS) analysis (see for an extensive survey Nijkamp, 1979). MDS methods can be used to transform ordinal data into cardinal results on the basis of two conditions: (1) a reduction of the original number of dimensions of the initial ordinal data matrix, and (2) a goodness-of-fit criterion which guarantees that the new cardinal configuration has a maximum correspondence with the original ordinal information. These MDS analyses have demonstrated their power in a large set of applications: perception and preference analysis, mental maps, recreation analysis, environmental quality analysis, urban renewal policy, and multicriteria evaluation methods (see also Voogd, 1978).

Finally, the use of logit models and more recently, probit models has to be mentioned (cf. Domenchich and McFadden, 1975, and Theil, 1971 b). These models are especially designed for analyzing discrete data (for instance, categorical data) on qualitative attributes of a certain phenomenon. These analyses are based on a probabilistic approach in which the frequencies of the occurrence of a phenomenon (or the shares of a variable in a whole set) are used as data input. Logit and probit analysis may especially be an important tool in an explanatory qualitative model, as will be set out in the next section.

Several ordinal data methods have - sometimes in combination with logit approaches - also been developed in the area of contingency table analysis (see Grizzle et al., 1969). In this respect, also chi-square approaches, dummy variable regression and analysis-of-variance approaches have been employed (see for applications among others Kuchler, 1978, Lehnen and Koch, 1974, and Margolin and Light, 1974).

### Soft Explanatory Models

Soft explanatory models are models which include soft variables either as dependent (response) variables or as explanatory (covariate) variables.

Following Wrigley (1979), one may make the following sub-division for the levels of measurement of covariate and response variables:

	response		
		cardinal	soft
covariate	cardinal	I	IV
	mixed	II	V
	soft	III	VI

Fig. 1. Levels of measurement of covariate and response variables.



A usual way of dealing with soft information (mainly nominal data) is the use of dummy variables. Dummy variables have often been used in statistical and econometric analyses in order to deal with non-metric explanatory variables such as occupational status, marital condition, sex, etc. (see Johnston, 1972, and Theil 1971 a). Dummy variables are numbers 0 and 1, that indicate whether or not a variable is a member of a given nominal class. They have mainly been used to include sociological and demographic variables in statistical and econometric analysis (see, among others, Orcutt et al., 1961); they belong to category II or III of Fig. 1.

An evident disadvantage of the dummy variable technique is that a situation with many distinct classes leads to a high number of dummy variables. Furthermore, a zero-one indicator for certain class characteristics does very often not use the available information in the most efficient way, as usually more than purely zero-one information is available (for instance, in a qualitative sense such as 'bad, normal, good'). In the latter case, more appropriate statistical and econometric methods do exist (see later).

Another way of including soft data in econometric models is the use of the above mentioned path models, as such models can be used to estimate causal or functional relationships between qualitatively-oriented clusters of variables. In this respect, especially the so-called Lisrel-models are important tools (see Jöreskog, 1977). The Lisrel procedure is based on a maximum likelihood approach; it needs accurate information about the distribution of the observed variables and the specification of the theoretical model. This approach belongs to class II or III of Fig. 1. Applications of the Lisrel approach can be found among others in Folmer (1980) and Jöreskog and Sörbom (1977).

A useful approach to soft econometrics is also provided by MDS methods (see Nijkamp, 1980). If the set of explanatory variables is composed of ordinal data (or includes a subset of ordinal data), then MDS analysis can be used to transform the ordinal information into cardinal information of a lower dimensionality. Next, this cardinal information can be used as data input in an explanatory model, so that then traditional estimation methods (such as least squares) may be applied. This approach bears a correspondence to a principal component analysis, which reduces a set of data into a subset of independent components, so that next these components can be included as covariate variables in a regression model. This MDS method also belongs to category II or III of Fig. 1.

Another class of soft econometric models is formed by the above mentioned logit and probit models. These stochastic models - based on observations on realized events of a discrete nature - may include various kinds of soft data (e.g., frequencies of occurrence of these events; cf. Theil, 1971 b, and Schmidt and Strauss, 1975 a, 1975 b).

Useful applications of logit models can be found among others in the field of categorical data analysis. Categorical data are discrete data (dichotomous or polychotomous data); they are often the result of survey questionnaires in which respondents have to indicate whether or not a certain attribute does exist or whether or not they regard this attribute as important. The frequency of positive or negative responses regarding these attributes can then be used as data input for the application of a logit model (see also Bishop et al., 1975, Theil, 1971 b, Upton, 1978). The variation in these frequencies (across the attributes) can then be related to a set of explanatory variables for these attributes (see for some applications also Koch and Reinfurt, 1971, and Wrigley, 1979, 1980 a, 1980 b). These models normally belong to class IV or V of Fig. 1.

Another application of logit analysis can be found in McCullagh (1980) who has developed a general class of regression models for ordinal data. These models are based on various modes of stochastic orderings of an ordinality structure. The author proposes the use of two models in particular, viz. the proportional odds model (based on a linear or non-linear logit model for the ordered categories of two response variables, given the values of covariates) and the proportional hazards model (based on a complementary log-log transformation of a hazard function for a response variable that depends on the difference between the covariates). Alternative approaches can be found among others in Landis and Koch (1977). These models may belong to class IV-VI of Fig.1.

A final way of dealing with ordinal information, both on the dependent and the explanatory variables, is found in Nijkamp and Rietveld (1981). This approach is based on a pairwise comparison of ordinal data followed by a dominance analysis via so-called regimes (sets of combinations of dominance and non-dominance relationships). Then the frequencies of such regimes can be included in a linear logit model, so that one may infer probability statements concerning the effect of a specific regime of covariate variables on the occurrence of the dominance or response variables. The latter class of models may fall into the categories IV - VI.

#### Discrepancy Analysis by means of Soft Statistics

This section will be devoted to a more detailed analysis of some statistical concepts and methods which may be relevant for discrepancy analysis. In each case first the cardinal variant will be discussed, followed by the ordinal analogon. Because of the limited space, only the main features of the methods will be presented. For further details on the methods we refer to Nijkamp and Rietveld (1981). The methods presented here will be applied in the last section of this paper.

The starting point of regional discrepancy is a vector of welfare components  $(x_{1r}, \dots, x_{Jr})$  of region  $r$ , a regional welfare profile. Elements of such a welfare profile may be socio-economic, environmental or infrastructure variables, etc. We will assume that all variables have been defined such that a higher value is preferred to a lower one. This means that variables for which small values are most attractive (e.g. unemployment) will be premultiplied by  $-1$ .

When one wants to characterize regional welfare level by means of one summary indicator, a regional welfare function

$$w_r = w_r(x_{1r}, \dots, x_{Jr})$$

has to be introduced. In general, it is not easy to specify such a function. In certain cases it is possible to draw conclusions about the relative welfare levels in various regions without any information about the specification of  $w_r$ . For example, when for all elements  $j$  of the welfare profile the performance in region 1 is better than in region 2, the welfare level in region 1 is better than in region 2, irrespective of the form of the welfare function.

The subject of regional discrepancy analysis is the data matrix  $X$

$$X = \begin{pmatrix} x_{11} & \dots & x_{1R} \\ \vdots & & \vdots \\ x_{J1} & \dots & x_{JR} \end{pmatrix} \quad (1)$$

formed by the welfare profiles of the  $R$  regions considered. The focus

will be on the interdependencies among the profile elements and on the (dis)similarities among the regions.

### 1. Correlation Analysis

The analysis of correlations between the profile elements is an important step in a regional discrepancy analysis. High correlations indicate that the set of profile elements can be reduced to a set of a more treatable size without loss of much information.

The standard measure to express the correlation between two cardinal variables  $x$  and  $y$  is the Pearson product-moment correlation-coefficient  $q$ :

$$q = \frac{\sum_r (x_r - \bar{x})(y_r - \bar{y})}{\sqrt{\sum_r (x_r - \bar{x})^2} \sqrt{\sum_r (y_r - \bar{y})^2}} \quad (2)$$

where  $\bar{x}$  and  $\bar{y}$  are the mean values of the  $x_r$  and  $y_r$ , respectively.

For ordinal data, Kendall's  $\tau_b$  is a well-known correlation measure (see Kendall, 1970):

$$\tau_b = \frac{s^+ - s^-}{\sqrt{s^+ + s^- + T_x} \sqrt{s^+ + s^- + T_y}} \quad (3)$$

where  $s^+$  and  $s^-$  are the number of concordant and discordant pairs of regions, respectively, and where  $T_x$  and  $T_y$  are the number of ties in  $x$  and  $y$ , respectively.

Consider the following operation for the ordinal data:

For all  $R(R-1)$  pairs  $r, s$ :

$$\begin{cases} x_{rs} = 1 & \text{if } x_r > x_s \\ x_{rs} = 0 & \text{if } x_r = x_s \\ x_{rs} = -1 & \text{if } x_r < x_s \end{cases} \quad (4)$$

The variable  $y_{rs}$  can be defined in the same way. Then it can be shown (see Hawkes, 1971 and Nijkamp and Rietveld, 1981), that

$$\tau_b = \frac{\sum x_{rs} y_{rs}}{\sqrt{\sum x_{rs}^2} \sqrt{\sum y_{rs}^2}} \quad (5)$$

which is basically the same analytical expression as (2).

This similarity result for simple correlation has given rise to several generalizations for multiple correlations (see Hawkes, 1971). One of the results is a coefficient of multiple correlation for ordinal data based on the above mentioned operation.

### 2. Principal Component Analysis

The aim of principal component analysis is the representation of  $J$  variables by a smaller number of variables (called components) with a high degree of accuracy. The aim is achieved by transforming the variables to a set of independent variables (based on an orthogonal data transformation in which the original variables are substituted for independent components).

In the case of ordinal data, components can be determined in various ways. One possibility is to make use of the fact that principal components can be expressed in terms of the correlation coefficients of the original variables.

Hence, with ordinal data, principal components can be found by employing rank order correlation coefficients such as  $\tau_b$ .

Another possibility is to determine a component such that the sum of the J rank correlation coefficients between this variable and the original variables is as large as possible. As indicated in Nijkamp and Rietveld (1981) there are various ways to give concrete form to this possibility (see also relationship (13)).

A drawback of principal component analysis (both for the ordinal and cardinal case) is the fact that new artificial variables are created which can be interpreted on the basis of factor loadings, but which have no clear direct meaning per se. In this respect, a more recently developed technique, called interdependence analysis, may be more appropriate. This will be the subject of the following paragraph.

### 3. Interdependence Analysis

Interdependence analysis is an optimal subset selection technique, by means of which a subset of variables which best represent an entire variable set can be chosen (see Boyce et al., 1974). The advantage of interdependence analysis is that an optimal subset of original variables is selected, so that a data transformation is not necessary.

Suppose we have matrix X with R observations (profiles) on J variables. Next, P variables are to be selected from the J variables such that this subset of P variables demonstrates an optimal correspondence with respect to the original data set. Hence, (J-P) variables are to be eliminated.

Now interdependence analysis is based on a series of successive regression analyses between the individual 'dependent' (J-P) variables to be eliminated and the 'independent' variables to be retained. Given (J-P) regressions, the minimum correlation coefficient can be calculated. Next, for all permutations of P in (J-P) variables, a similar operation can be carried out. Then the optimal subset is defined as that subset which maximizes over all permutations the values of the above-mentioned minimum correlation coefficient.

In the context of ordinal data, interdependence analysis is feasible, once a rank order multiple correlation coefficient is available. As indicated in par.1, such a coefficient can be developed, so that we may conclude that ordinal interdependence analysis is indeed feasible.

### 4. Inequality Measures

Inequality measures indicate the intensity of the inequality among a series of observations (regions). These measures are, in general, defined in terms of the distances of the observations to a reference point (e.g. the mean). Inequality measures are standardized: they are independent from the unit of measurement. A well-known example of an inequality measure is the coefficient of variation which is defined as the ratio of the standard deviation and the mean of a distribution:

$$v = \frac{\sqrt{\frac{1}{R} \sum (x_r - \bar{x})^2}}{\bar{x}} \quad (6)$$

In Blommestein et al. (1981) a generalization of inequality measures for multidimensional welfare profiles is contained.

In employing ordinal data, the concept of an inequality measure is problematic, since the central elements of an inequality measure - magnitudes of difference with respect to a reference point - are not contained in ordinal data. Yet, there is some scope for using equality measures with ordinal data.

Consider the following ordinal data matrices for 4 regions and 2 variables:

$$X_1 = \begin{pmatrix} 1 & 4 & 3 & 2 \\ 1 & 4 & 3 & 2 \end{pmatrix} \quad X_2 = \begin{pmatrix} 1 & 4 & 3 & 2 \\ 3 & 1 & 2 & 4 \end{pmatrix} \quad (7)$$

Applying equality measures to the sum of the two scores in  $X_1$  and  $X_2$  indicates a larger inequality in  $X_1$  than in  $X_2$ . This result reflects that in  $X_2$  the inequality in variable 1 is to a certain extent compensated by the equality in variable 2, while in  $X_1$  the inequality in the variables works in the same direction.

Given the nature of ordinal data, the summation of rank orders as suggested above, is questionable, however. A better way of measuring the extent to which welfare inequality is increasing or decreasing by taking simultaneously into account various welfare components is the following one.

Let  $J_{rs}^+$  be the number of variables for which region  $r$  is preferred to  $s$  and  $J_{rs}^-$  the number of variables for which region  $s$  is preferred to  $r$ . Then  $\sum_{r,s} (J_{rs}^+ - J_{rs}^-)$  is a measure of the extent to which a low value for a certain variable occurs simultaneously with low values for other variables. After standardization we arrive at the measure:

$$\gamma = \frac{\sum_{r,s} |J_{rs}^+ - J_{rs}^-|}{\sum_{r,s} (J_{rs}^+ + J_{rs}^-)} \quad (8)$$

so that  $0 \leq \gamma \leq 1$ . When (8) is applied to  $X_1$  and  $X_2$  in (7), one arrives at values 1 and 1/6, respectively, which is in accordance with the conclusion that in  $X_2$  inequality compensation takes place. When the number of variables is odd, a correction has to be applied to (8), since then  $|J_{rs}^+ - J_{rs}^-|$  cannot adopt the value zero (assuming that no ties occur).

## 5. Cluster Analysis

The aim of cluster analysis is the derivation of sets of individuals (regions) or variables which are in a certain sense similar. There are many types of clustering methods (see Hartigan, 1975). Clustering methods can be distinguished, among others, according to:

- the similarity measure (e.g., the correlation coefficient between two variables)
- the objective function (e.g., the objective may be: maximize the similarity within clusters, minimize the similarity between clusters)
- the way in which clusters are formed (hierarchical versus non-hierarchical).

In the case of ordinal data the main problem is the construction of a suitable similarity criterion. When the aim is a clustering of variables, one can simply employ a rank correlation coefficient as a similarity measure.

When the aim is a clustering of regions according to welfare levels, a similarity measure can be employed that is also used in multicriteria analysis (cf. Rietveld, 1980). This measure can

be developed as follows. The term  $(J_{rs}^+ - J_{rs}^-)$  defined in par.4 denotes the net number of variables according to which region r is better than region s. Hence  $(J_{rs}^+ - J_{rs}^-)$  can be conceived of as distance between r and s.

Then

$$\xi = J - (J_{rs}^+ - J_{rs}^-) \quad (9)$$

is a similarity measure for r and s which is suitable for ordinal data. Once a similarity measure is given, the application of cluster analysis to ordinal data can be carried out along the same lines as with cardinal data. Note, that an implicit assumption underlying (9) is that all variables are equally important.

#### Discrepancy Analysis by means of Soft Econometrics

Explanatory models in a regional discrepancy analysis may focus on two different subjects. A first kind of analysis may focus on the causes of spatial discrepancies, for instance, lack of infrastructure facilities. A second class of models may focus on the consequences of spatial discrepancies, for instance, migration movements. In this respect, regional welfare inequalities may play a central role in investment and location decisions of firms and households. Hence, when explaining interregional migration, the components of the regional welfare profiles have to be taken into account as explanatory variables. Clearly, when cardinal data on the welfare profiles are available, there is no need for soft econometric methods. But in any other case, soft econometric explanations models may be very meaningful.

##### 1. Generalized Rank Correlation Analysis

This approach to ordinal data has already been indicated in paragraphs 1, 2 and 3 of the preceding section. The main idea is that the ordinal data matrix X of size (J x R) and the vector of dependent variables Y of size (1 x R) are transformed into a matrix  $\bar{X}$  of size (J x R (R-1)) and a vector  $\bar{Y}$  of size (1 x R (R-1)) by means of procedure (4). The matrix and vector thus derived consist of values 1, 0 and -1. Then  $\bar{X}$  and  $\bar{Y}$  can serve as inputs for the usual least squares procedures.

##### 2. Logit Analysis

Consider the following linear relationship:

$$y_r = \alpha_0 + \alpha_1 x_{1r} + \dots + \alpha_J x_{Jr} \quad (r = 1, \dots, R) \quad (10)$$

When (10) is rewritten in terms of differences, we get the following result:

$$y_r - y_s = \alpha_1 (x_{1r} - x_{1s}) + \dots + \alpha_J (x_{Jr} - x_{Js}) \\ (r, s = 1, \dots, R \quad r \neq s) \quad (11)$$

When transformation (4) is applied to the R(R-1) pairs of observations in (11), each pair of observations on the independent variables is characterized by a J vector consisting of values 1, 0 and -1. Such a vector will be called a regime. In principle there are  $3^J$  different regimes.

Consider a particular regime m, and let  $F_m$  be the number of pairs of observations giving rise to regime m. For these pairs of observations there are three possibilities:  $y_r$  may be larger than, equal to, or less than  $y_s$ . The number of these pairs will be denoted as  $F_{m1}$ ,  $F_{m0}$ , and  $F_{m-1}$ , respectively. Thus  $f_{m1} = F_{m1}/F_m$  is the probability that, given that regime m holds for a pair of observations (r, s), the value of  $y_r$  is larger than  $y_s$ . In the same way,  $f_{m0}$  and  $f_{m-1}$  can be defined.

The following relationship relates the values of  $f_{m1}$  and  $f_{m-1}$  to the structure of the corresponding regime:

$$\ln \left( \frac{f_{m1}}{f_{m-1}} \right) = \beta_0 + \beta_1 z_{m1} + \dots + \beta_J z_{mJ} \quad (12)$$

where  $(z_{m1}, \dots, z_{mJ})$  is a series of values 1, 0 and -1 characterizing regime  $m$ . For details on the reason of the logit specification in (12) and the way to estimate the parameters  $\beta_j$  we refer to Theil (1971b) and Nijkamp and Rietveld (1981).

#### A Multivariate Analysis of Dutch Interregional Welfare Discrepancies

The concepts and methods presented in the preceding sections will be applied to Dutch regional data<sup>1)</sup> from 1976 - '78. Data have been collected for 40 regions and 13 profile elements ( $R = 40$ ,  $J = 13$ ). The location of the regions has been depicted in Fig. 2. The mean population size of the regions is approximately 350.000 inhabitants. The socio-economic variables are:

1. fiscal income per capita
2. unemployment rate
3. wealth per capita
4. index of cost of living.

The environmental variables are:

5. population density
6. size of natural environment. as percentage of total regional area
7. index of industrialization related to regional area
8. index of the emission of pollutants related to regional area.

The infrastructural variables are:

9. density of transport network
10. index of cultural centres and sport accommodations per capita
11. index of the number of schools of various types per capita
12. distance to the centre of the Netherlands
13. index of various medical services per capita.

For a precise definition and presentation of the variables as well as the sources of the data, we refer to Van Veenendaal (1981).

The variables 2, 4, 5, 7, 8 and 12 have been multiplied with a factor -1 so that for all variables a larger value is preferred to a smaller one.

In the following, we will first present the results of a cardinal analysis, followed by an ordinal analysis, based on the same data matrix.

1) The authors thank Wouter van Veenendaal and Jan Broersma for their assistance in the construction of the data set.



Fig. 2. Location of 40 regions in the Netherlands.



### 1 a. Correlation Analysis: Cardinal

The correlation coefficients between the 13 variables have been represented in Table 1.

1	2	3	4	5	6	7	8	9	10	11	12	13
1.00	.57	.60	-.17	-.84	.30	-.53	-.83	.77	.05	.15	.62	.55
.57	1.00	.49	-.13	-.40	.31	-.54	-.37	.25	-.15	.17	.72	.24
.60	.49	1.00	.06	-.39	.29	-.11	-.36	.36	.12	.15	.28	.46
-.17	-.13	.06	1.00	.04	-.24	-.09	.01	-.01	.17	-.06	-.28	.04
-.84	-.40	-.39	.04	1.00	-.19	.75	.97	-.94	-.10	-.01	-.52	-.47
.30	.31	.29	-.24	-.19	1.00	-.17	-.09	.15	-.04	.04	.25	.19
-.53	-.36	-.11	-.09	.75	-.17	1.00	.89	-.68	-.03	.03	-.33	-.40
-.83	-.37	-.36	.01	.97	-.09	.89	1.00	-.91	-.11	-.01	-.52	-.47
.77	.25	.36	-.01	-.94	.15	-.68	-.91	1.00	.04	-.02	.46	.38
.05	-.15	.12	-.17	-.10	-.04	-.03	-.11	.04	1.00	.01	-.17	.09
.15	.17	.15	-.06	-.01	.04	.03	-.01	-.02	.01	1.00	.12	.27
.62	.72	.28	-.28	-.52	.25	-.33	-.62	.46	-.17	.12	1.00	.18
.55	.24	.46	.04	-.47	.19	-.40	-.47	.38	.09	.27	.18	1.00

Table 1. Correlation matrix for cardinal data.

Within the three blocks of variables the correlations are predominantly positive. Relatively high correlations can be found in the socio-economic profile (variables 1-3) and the environmental profile (variables 5, 7 and 8). The correlations between the infrastructural variables are relatively small. We may conclude, therefore that the homogeneity of the variables in the socio-economic and environmental profiles is clearly larger than in the infrastructural profile.

A further inspection of the table teaches that in the set of 13 variables two clearly different subsets of variables  $s_1$  and  $s_2$  can be distinguished.

The first subset consists of the variables 5, 7 and 8, the other subset consists of the variables 1, 2, 3, 6, 9, 12 and 13. All correlations between variables of the same subset are positive, while the correlations between variables of different subsets are negative. The variables of the first subset are clearly related to the negative aspects of urbanized regions, while the variables of the second subset are mainly related to the positive aspects of these regions (the occurrence of variable 6 in this subset is somewhat astonishing). This suggests that a considerable part of interregional welfare differences is related with the level of urbanization. There are only three variables (viz. 4, 10 and 11) which cannot be classified in one of the subsets. For the variables 10 and 11, this may reflect that the governmental policies in the provision of cultural, recreational and educational facilities have been more or less neutral with respect to the level of urbanization.

### 1 b. Correlation Analysis: Ordinal

The rank correlation coefficients between the variables have been represented in Table 2.

1	2	3	4	5	6	7	8	9	10	11	12	13
1.00	.44	.25	-.19	-.56	.13	-.48	-.57	.50	-.07	-.08	.55	.40
.44	1.00	.37	-.06	-.38	.11	-.24	-.34	.25	-.16	-.12	.59	.21
.25	.37	1.00	-.04	-.64	.02	-.01	-.10	.02	.03	.13	.19	.28
-.19	-.06	-.04	1.00	.16	-.14	.06	.13	-.19	.10	-.01	-.19	-.13
-.56	-.38	-.64	.16	1.00	-.11	.74	.82	-.74	.11	.04	-.53	-.43
.13	.11	.02	-.14	-.11	1.00	-.06	-.05	.10	.07	.10	.09	.24
-.48	-.24	-.01	.06	.74	-.06	1.00	.70	-.67	.03	-.01	-.36	-.45
-.57	-.34	-.10	.13	.82	-.05	.70	1.00	-.72	.13	.06	-.57	-.40
.50	.25	.02	-.19	-.74	.10	-.67	-.72	1.00	-.13	-.07	.50	.33
-.07	-.16	.03	.10	.11	.07	.03	.13	-.13	1.00	.37	-.23	.07
-.08	-.12	.13	-.01	.04	.10	-.01	.06	-.07	.37	1.00	-.21	.20
.55	.59	.19	-.19	-.53	.09	-.36	-.57	.50	-.23	-.21	1.00	.21
.40	.21	.28	-.13	-.43	.24	-.45	-.40	.33	.07	.20	.21	1.00

Table 2. Correlation coefficients for ordinal data.

When we compare Table 1 with Table 2, we find that the order of magnitude of the coefficients in the first table is somewhat larger than in the second one. This is a common result when comparing product-moment and rank correlation coefficients. The sign of the correlations is in most cases the same in both tables.

The conclusion, based on Table 1, that the homogeneity of the variables in the socio-economic and environmental profiles is clearly larger than in the infrastructure profile cannot be maintained for Table 2. Table 2 gives rise to the conclusion that the homogeneity of the infrastructure and socio-economic variables is more or less the same, but clearly smaller than in the environmental profile.

The other conclusion, based on Table 1, that two clearly different subsets of variables can be distinguished, can be maintained for Table 2 with almost the same subsets:

$$s_1 = \{ 4, 5, 7, 8 \} \text{ and } s_2 = \{ 1, 2, 3, 6, 9, 12 \}.$$

## 2 a. Principal Component Analysis: Cardinal

We will focus the attention on the first principal component. This component accounts for 42 % of the total variation in the data matrix. Its loadings have been represented in the first row of table 3. The loadings follow exactly the pattern

variable	1	2	3	4	5	6	7	8	9	10	11	12	13
cardinal data	.94	.61	.57	-.10	-.93	.32	-.73	-.90	.86	.04	.13	.67	.59
ordinal data	.76	.54	.23	-.23	-.87	.17	-.77	-.87	.80	-.17	-.09	.72	.54

Table 3. Loadings of the first principal component.

discovered in the correlation analysis: strong negative values for the elements of  $s_1$ , strong positive values for the elements of  $s_2$  and values near to zero for the remaining variables. Hence it is reasonable to interpret the first principal component as an indicator of urbanization, which is positively related to the main socio-economic variables and negatively to the main environmental variables.

## 2 b. Principal Component Analysis: Ordinal

The first principal component has been found by applying an ordinary principal component analysis to the data matrix obtained by transformation (4). It accounts for 36 % of the total variation in the transformed data matrix. Its loadings can be found in the second row of Table 3. The similarity with the loadings for cardinal data is striking. We may conclude that for the first component, ordinal data give rise to virtually the same results as cardinal data. We have checked whether this conclusion also holds true for the second and third principal component (each accounting for approximately 12 % of the total variation). The answer is negative: there is no clear correspondence between the loadings of these components for the cardinal and the ordinal case.

A second type of ordinal principal component analysis, aiming at the derivation of a summary indicator of regional welfare, has been carried out as follows. The ordinal principal component is defined as the rank order of the regions which is the solution of:

$$\max! \quad \frac{1}{J} \sum_j \tau_b^j \quad (13)$$

where  $\tau_b^j$  indicates the rank correlation coefficient between the principal component and variable  $j$ . Obviously this principal component is measured on an ordinal scale. A principal component has been computed for the welfare profile as a whole and also for the three subprofiles separately (see Table 4). High rank orders for the total welfare level are obtained mainly in the western and central part of the Netherlands. Low rank orders occur in the peripheral regions. Obviously, these results depend on the implicit assumption underlying (13) that all variables are equally important.

region	principal component (ordinal)	unweighted average (cardinal)	region	principal component (ordinal)	unweighted average (cardinal)
1	6.	.372	21	40.	.646
2	5.	.353	22	11.	.432
3	19.	.423	23	34.	.523
4	20.	.434	24	39.	.643
5	27.	.493	25	17.	.515
6	32.	.447	26	37.	.580
7	30.	.486	27	8.	.435
8	12.	.373	28	10.	.424
9	33.	.475	29	23.	.479
10	28.	.466	30	9.	.451
11	35.	.498	31	21.	.470
12	18.	.462	32	31.	.502
13	29.	.588	33	13.	.461
14	24.	.480	34	14.	.471
15	26.	.505	35	7.	.434
16	4.	.446	36	15.	.459
17	38.	.644	37	3.	.429
18	16.	.488	38	2.	.401
19	36.	.524	39	1.	.344
20	25.	.479	40	22.	.455

Table 4. Aggregate welfare levels for 40 regions based on ordinal and cardinal data.

When information is available about the weights to be attached to the variables, (13) can easily be reformulated as the maximization of a weighted average of rank correlation coefficients.

Table 4 also contains the unweighted average value of the 13 variables (after an appropriate standardization) based on the cardinal data. The rank order coefficient between these variables is .64, which indicates that these variables are to a reasonable extent in agreement with each other.

### 3.a. Interdependence Analysis : Cardinal

The results of a cardinal interdependence analysis have been represented in Table 5. In the last column of the table, the minimum multiple correlation coefficient resulting from the regression of each

Number of Selected Variables	Selected Variables	Minimum Multiple Correlation Coefficient
1	2	.02
2	12, 13	.04
3	8, 12, 13	.08
4	6, 10, 11, 12	.13

Table 5. Interdependence analysis (cardinal)

variable with the selected variables as independent variables has been given. The set of selected variables is not very stable. Only variable 12 is a stable member of this set. This means that the composition of the set of selected variables depends to a considerable extent on the number of variables selected, which is not a satisfactory result.

### 3.b. Interdependence Analysis : Ordinal

The results of an ordinal interdependence analysis have been represented in Table 6. Also for ordinal variables, a rather unstable pattern of selected variables arises. When we compare the results of

Number of Selected Variables	Selected Variables	Minimum Multiple Correlation Coefficient
1	12	.01
2	12, 13	.04
3	4, 10, 13	.07
4	1, 4, 6, 11	.09

Table 6. Interdependence Analysis (ordinal)

an ordinal and a cardinal analysis, we conclude that there are similarities (see for example the case of two selected variables), but that in most cases the patterns of selected variables are certainly not identical.

#### 4 a. Inequality Measures: Cardinal

A well-known inequality measure is the coefficient of variation V. This inequality measure has been computed for 13 variables. The result is shown in Table 7. The table indicates that the interregional inequality in the socio-economic variables

var.	1	2	3	4	5	6	7	8	9	10	11	12	13
V	.08	.41	.34	.07	.88	.70	.76	.94	.42	.44	.17	1.24	.43

Table 7. Values of the coefficients of variation V for 13 variables.

is relatively small, while the inequality in the environmental variables is relatively large. For the infrastructural variables we find in most cases intermediate positions.

The above statements hold for the variables in the three sub-profiles independently. It is also interesting to know the degree of inequality for a composite variable representing a whole sub-profile. These composite variables have been constructed by calculating the unweighted average of the normalized variables in each profile. For the three composite variables, we find as outcomes for V : (.17, .63, .31). When we compare this outcome with the mean values of V in Table 7 per sub-profile: (.22, .82, .54), we note that the rank order is the same and that in all cases the inequality in the composite variable is smaller than in the constituent variables. The relative and absolute decrease is largest for the infrastructure variable, which means that inequality compensation occurs to a larger extent in this sub-profile than in the other sub-profiles.

#### 4 b. Inequality Measures: Ordinal

As indicated in the paper, in case of ordinal data the phenomenon of inequality compensation can be analyzed by means of the measure  $\gamma$ , as defined in (8). For the three sub-profiles we arrive at the following values for  $\gamma$  : (.49, .64, .29). This means that inequality compensation occurs to a larger extent in the socio-economic and environmental sub-profile than in the infrastructure sub-profile. This conclusion is in accordance with the conclusion of the cardinal analysis.

#### 5 a. Cluster Analysis: Cardinal

We will present the results of analyzing clusters of variables. As similarity measures, the correlation coefficients, displayed in Table 1, have been adopted. The way of combining clusters is non-hierarchical. The clustering objective is the maximization of the internal homogeneity of the clusters, measured as the minimum correlation between any two variables in one of the clusters. Table 8 shows the outcomes.

# clusters	composition of clusters	minimum correlation between cluster elements
4	$C_1 = \{ 1, 2, 3, 6, 9, 12, 13 \}$ $C_2 = \{ 5, 7, 8 \}$ $C_3 = \{ 4, 10 \}$ $C_4 = \{ 11 \}$	$r_{6,9} = .154$
3	$C_1 = \{ 1, 2, 3, 6, 9, 12, 13 \}$ $C_2 = \{ 5, 7, 8, 11 \}$ $C_3 = \{ 4, 10 \}$	$r_{5,11} = -.015$
2	$C_1 = \{ 1, 2, 3, 6, 9, 12, 13 \}$ $C_2 = \{ 4, 5, 7, 8, 10, 11 \}$	$r_{8,10} = -.109$
1	$C_1 = \{ 1, 2, \dots, 13 \}$	$r_{5,9} = -.941$

Table 8. Cluster analysis of 13 variables (cardinal data).

When the aim is the formation of clusters with exclusively positive internal correlations, at least four clusters have to be created. A large jump occurs in the objective function with the transition from 2 to 1 cluster: a decrease from  $-.109$  to  $-.941$ . This indicates that in the data set two main classes of variables can be distinguished.

#### 5 b. Cluster Analysis: Ordinal

A cluster analysis has been carried out along the same lines as in 5.a, based on the rank correlation coefficients of Table 2. The results are displayed in Table 9. The table shows that

# clusters	composition of clusters	minimum correlation between cluster elements
4	$C_1 = \{ 1, 2, 6, 9, 12, 13 \}$ $C_2 = \{ 4, 5, 7, 8 \}$ $C_3 = \{ 3 \}$ $C_4 = \{ 10, 11 \}$	$r_{4,7} = .059$
3	$C_1 = \{ 1, 2, 6, 9, 12, 13 \}$ $C_2 = \{ 4, 5, 7, 8 \}$ $C_3 = \{ 3, 10, 11 \}$	$r_{3,10} = .028$
2	$C_1 = \{ 1, 2, 3, 6, 9, 12, 13 \}$ $C_2 = \{ 4, 5, 7, 8, 10, 11 \}$	$r_{4,11} = -.007$
1	$C_1 = \{ 1, 2, \dots, 13 \}$	$r_{5,9} = -.737$

Tabel 9. Cluster analysis of 13 variables (ordinal data).

at least three clusters have to be created when one aims at forming clusters with exclusively positive correlations. The differences between a cardinal and an ordinal analysis are small with this data set.

# 6 a. Interregional Migration and Interregional Welfare Discrepancies: Cardinal Analysis

In this paragraph we will examine the extent to which inter-regional migration can be explained by interregional welfare discrepancies. The following equation will be the subject of estimation:

$$\frac{M^{rs} - M^{sr}}{P^r \cdot P^s} = (d^{rs})^{-\alpha_0} \{ \alpha_1 (y_1^r - y_1^s) + \dots + \alpha_4 (y_4^r - y_4^s) \} \quad (14)$$

where  $M^{rs}$  : number of migrants per year from r to s  
 $P^r$  : population of region r  
 $d^{rs}$  : distance between r and s  
 $y_k^r$  : feature k of region r ( $k = 1, \dots, 4$ ).

The variables  $y_1$ ,  $y_2$  and  $y_3$  in the analysis are summary indicators for the three sub-profiles, obtained by an unweighted summation of the normalized variables. In order to include the situation on the housing market as an explanatory variable for migration, we added as a fourth regional feature: the availability of dwellings. This variable is measured as the difference between the rate of change of the housing stock and the rate of change of population (apart from migration) For a further discussion of equation (14) we refer to Suyker (1980).

Equation (14) gives rise to 5 parameters to be estimated, given 780 ( $\frac{1}{2} \cdot 40 \cdot 39$ ) observations. We estimated  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ , given various predetermined values of  $\alpha_0$ . The best results have been obtained when  $\alpha_0 = 1.3$  (see Table 10)

$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$R^2$
1.30	-1.25 (1.47)	3.42 (2.66)	-.68 (-.76)	3.69 (15.08)	.296

Table 10. Estimation results of the migration relationship  
(t-values between brackets).

The estimations of  $\alpha_2$  and  $\alpha_4$  (environment and housing) have the right sign and are significantly different from zero at the 1% level. The estimations of  $\alpha_1$  and  $\alpha_3$  (socio-economic conditions and infrastructure) are negative, though not at a significant level. Since good socio-economic conditions and infrastructure contribute to regional welfare, and hence to the attractiveness of a region as a destination of migration, the negative coefficients are unsatisfactory, though not completely unexpected. The migration pattern in the Netherlands is characterized by a strong outmigration from the highly urbanized regions towards other regions. The urban regions have in general a favourable socio-economic and infrastructural position. Negative signs for  $\alpha_1$  and  $\alpha_3$  may indicate that  $y_1$  and  $y_3$  serve as proxy variables for the negative aspects of urbanized regions. This would mean that  $y_2$  and  $y_4$  do not exhaustively represent these negative aspects. Especially for  $y_4$ , this is not a surprising conclusion, since (due to data limitations) it has been defined in a rather crude way.

6 b. Interregional Migration and Interregional Welfare Discrepancies:  
Ordinal Analysis

The ordinal analogue of (14) has been estimated as follows:

1. Solve (13) for each of the three sub-profiles in order to derive a summary indicator  $y_k$  for each of these profiles ( $k = 1, 2, 3$ ).
2. Apply transformation (4) to (14). The result is a series of 16 different regimes with accompanying logits. See Table 11.

regime	regime structure	F	$F_1$	$F_{-1}$	$F_1/F$	$\ln \left( \frac{F_1}{F_{-1}} \right)$
1	-1 -1 -1 -1	46	16	29	.35	-.59
2	1 -1 -1 -1	53	7	45	.13	-1.86
3	-1 1 -1 -1	140	106	33	.76	1.17
4	1 1 -1 -1	87	45	42	.52	.07
5	-1 -1 1 -1	90	30	59	.33	-.68
6	1 -1 1 -1	258	38	217	.15	-1.74
7	-1 1 1 -1	56	33	20	.59	.59
8	1 1 1 -1	49	22	27	.45	-.20
9	-1 -1 -1 1	49	27	22	.55	.29
10	1 -1 -1 1	56	20	33	.36	-.58
11	-1 1 -1 1	258	217	38	.84	1.74
12	1 1 -1 1	90	59	30	.66	.68
13	-1 -1 1 1	87	42	45	.48	-.07
14	1 -1 1 1	140	33	106	.24	-1.17
15	-1 1 1 1	53	45	7	.85	1.86
16	1 1 1 1	46	29	16	.63	.59

Table 11. Regimes used in the logit analyses for ordinal data.

The left part of the table shows the structure of the regimes. The regimes with the largest frequency are (1 -1 1 -1) and its complement (-1 1 -1 1), which indicates the negative correlations between  $y_1$  and  $y_3$  on the one hand and  $y_2$  and  $y_4$  on the other hand.

There are only some pairs of regions for which migration in both directions is equal, which would mean that net migration is zero. Therefore we find for all regimes that the sum of  $F_1$  and  $F_{-1}$  is near or equal to F.

The last column of the table shows the logit expression: it is positive when  $F_1 > F_{-1}$  and negative when  $F_1 < F_{-1}$ .



A weighted least squares estimation of (12) based on the data of Table 11 gives rise to the following results (see Table 12).

$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$R^2$
-.51	.78	-.14	.36	.986

Table 12. Estimation of the migration relationship for ordinal data.

The signs of the coefficients are the same as with cardinal data. The infrastructure coefficient is in both experiments clearly smaller (in absolute sense) than the other coefficients. Given the large interdependencies in Table 11, it is not meaningful to compute t-values.

### Conclusion

Concerning the results of an ordinal and cardinal analysis of regional discrepancies, we conclude that for many methods the results are more or less equal, although in some cases the results of a cardinal analysis are more powerful than those of an ordinal one (for example in the field of inequality measures (par.4) and the significance of estimated coefficients (par.6). No clear contradictions have been found between a cardinal and ordinal analysis, which means that ordinal methods are in most respects a useful tool for discrepancy analysis. Consequently, an ordinal level of measurement does not preclude a meaningful inference of statistical and econometric conclusions.

## References

- Adelman, I., and C.T. Morris, The Deviation of Cardinal Scales from Ordinal Data, Economic Development and Planning, (W. Sellekaerts, Ed.), MacMillan, London, 1974, pp. 1-39.
- Apel, H., Soft Models of Herman Wold's Type, Paper Fourth Advanced Summer Institute in Regional Science, Siegen, 1978 (mimeographed).
- Bishop, Y.M.M., S.E. Fienberg and P.W. Holland, Discrete Multivariate Analysis, MIT Press, Cambridge, Mass., 1975.
- Blalock, H.M., Can We Find a Genuine Ordinal Scope Analogue ?, Sociological Methodology (D.R. Heise, Ed.), Jossey-Bass Publ., San Francisco, 1976, pp. 195-226.
- Blalock, H.M., Causal Inferences in Nonexperimental Research, University of North Carolina Press, Chapel Hill, 1964.
- Blommestein, H.J., P. Nijkamp and P. Rietveld, A Multivariate Analysis of Spatial Inequalities, in W. Buhr and P. Friedrich (eds.), Regional Development under Stagnation, Nomos, Baden-Baden, 1981.
- Boyce, D.E., A. Fahbi and R. Weischedel, Optimal Subset Selection, Springer, Berlin, 1974.
- McCullagh, P., Regression Models for Ordinal Data, Journal of the Royal Statistical Society, vol. 42, no. 2, 1980, pp. 109-142.
- Domencich, F.A., and D. McFadden, Urban Travel Demand, a Behavioral Analysis, North Holland Publ. Co., Amsterdam, 1975.
- Folmer, H., Measurement of the Effects of Regional Economic Policy Instruments, Environment and Planning, vol. , 1980, pp.
- Grizzle, J.E., C.F. Starmer and G.G. Koch, Analysis of Categorical Data by Linear Models, Biometrics, vol. 25, 1969, pp. 489-504.
- Hartigan, J.A., Clustering Algorithms, Wiley, New York, 1975.
- Hawkes, R.K., The Multivariate Analysis of Ordinal Measures, American Journal of Sociology, vol. 76, 1971, pp. 908-926.
- Hui, B.S., Soft Modelling by the Partial Least Squares Approach, Ph. D. Diss., University of Pennsylvania, 1978.
- Johnston, J., Econometric Methods, McGraw-Hill, New York, 1972.
- Jöreskog, K.G., Structural Equation Models in the Social Sciences, Application of Statistics (P.R. Krishnaiah, Ed.) North Holland Publ. Co., Amsterdam, 1977, pp. 265-286.
- Jöreskog, K.G., and D. Sörbom, Statistical Models and Methods for Analysis of Longitudinal Data, Latent Variables in Socioeconomic Models (D.J. Aigner and A.S. Goldberger, Eds.) North-Holland Publ. Co., Amsterdam, 1977, pp. 285-325.
- Kendall, M.G., Rank Correlation Methods, Griffin, London, 1970.
- Koch, G.G. and D.W. Reinfurt, The Analysis of Categorical Data from mixed Models, Biometrics, vol. 27, 1971, pp. 157-173.
- Küchler, M., Alternativen in der Kreuztabellenanalyse, Zeitschrift für Soziologie, vol. 7, no. 4, 1978, pp. 347-365.

- Landis, J.R. and G.G. Koch, The Measurement of Observer Agreement for Categorical Data, Biometrics, vol. 33, 1977, pp. 159-174.
- Lehnen, R.G. and G.G. Koch, A general Linear Approach to the Analysis of Nonmetric Data, American Journal of Political Science, vol. 18, 1974, pp. 283-313.
- Margolin, B.H., and R.J. Light, An Analysis of Variance for Categorical Data, Journal of the American Statistical Association, vol. 69, 1974 pp. 755-764.
- Namboodiri, N.K., L.F. Carter and H.M. Blalock, Applied Multivariate Analysis and Experimental Designs, McGraw-Hill, New York, 1975.
- Nijkamp, P., Soft Econometric Models, Research Memorandum 1980-5, Dept of Economics, Free University Amsterdam, 1980.
- Nijkamp, P. and P. Rietveld, Ordinal Multivariate Analysis, Professional Paper PP-81-2, IIASA, Laxenbuy, 1981.
- Nijkamp, P., Multidimensional Spatial Data and Decision Analysis, Wiley, Chichester/New York, 1979.
- Orcutt, G.H., M. Greenberger, J. Korbel and A.M. Rivlin, Microanalysis of Socioeconomic Systems, Harper & Row, New York, 1961.
- Ploch, D.R. Ordinal Measures of Association and the General Linear Model, Measurement in the Social Sciences (H.M. Blalock, Ed.), MacMillan, London, 1974, pp. 343-368.
- Ponsard, C., On the Imprecision of Consumer's Spatial Preferences, Papers of the Regional Science Association, vol. 42, 1979, pp. 59-71.
- Ryans, A. B. and V. Srinivasan, Improved Method of Comparing Rank-Order Preferences of Two Groups of Consumers, Journal of Marketing Research, vol. 16, 1979, pp. 583-587.
- Rietveld, P., Multiple Objective Decision Methods and Regional Planning, North Holland Publ. Comp., Amsterdam, 1980.
- Schmidt, P. and R.P. Strauss, Estimation of Models with Jointly Dependant Qualitative Variables: A Simultaneous Logit Approach, Econometrica, vol. 43, no.4, 1975 a, pp. 745-755.
- Schmidt, P. and R.P. Strauss, The Prediction of Occupation Using Multiple Logit Models, International Economic Review, vol. 16, no. 2, 1975 b, pp. 471-486.
- Siegel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York, 1956.
- Suyker, W.B.C., An Analysis of Interprovincial Migration in the Netherlands, Paper presented on the Study Institute on Urban Problems and Policies in a Spatial Context, Soesterberg, 1980.
- Theil, H., Principles of Econometrics, Wiley, New York, 1971 a.
- Theil, H., On the Estimation of Relationships Involving Qualitative Variables, American Journal of Sociology, vol. 76, 1971 b, pp. 103-154.
- Upton, G.J.G., The Analysis of Cross-Tabulated Data, Wiley, New York, 1978.

- Veenendaal, W.M. van, Regionale Welvaart in Nederland (mimeographed), Department of Economics, Free University, Amsterdam, 1981.
- Voogd, J.H., On the Principles of Ordinal Geometric Scaling, Research Paper no. 10, Research Centre for Physical Planning PSC-TNO, Delft, 1978.
- Whalen, T., Risk Minimization Using L-Fuzzy Sets, Proceedings IEEE Conference on Cybernetics and Society, Cambridge, U.S.A., 1980, pp. 821-825.
- Wold, H., Soft Modeling by Latent Variables, Perspectives in Probability and Statistics (J. Gassi, Ed.) Academic Press, London, 1975, pp. 117-142.
- Wold, H., On the Transition from Pattern Recognition to Model Building, Mathematical Economics and Game Theory (R. Henn and O. Moeschlin, Eds.) Springer, Berlin, 1977, pp. 536-549.
- Wold, H., Model Construction and Evaluation when Theoretical Knowledge is Scarce, Cahier 79-06, Dept. of Econometrics, University of Geneva, 1979 (mimeographed).
- Wrigley, N., Development in the Statistical Analysis of Categorical Data, Progress in Human Geography, vol. 3, 1979, pp. 315-355.
- Wrigley, N., Categorical Data, Repeated-Measurement Research Designs, and Regional Industrial Surveys, Regional Studies, vol. 14, 1980 a, pp. 455-471.
- Wrigley, N., Paired Comparison Experiments and Logit Models, Environment & Planning A, vol. 12, 1980 b, pp. 21-40.
- Zadeh, L., Outline of a New Approach to the Analysis of Complex Systems and Decision Processes, Multiple Criteria Decision-Making (J.L. Cochrane and M. Zeleny, Eds.), University of South Carolina Press, South Carolina, 1973, pp. 686-725.

Serie Research Memoranda:

- |        |   |  |
|--------|---|--|
| 1979-1 | A.A. Schreuder-Sunderman,<br>Frans Blommaert en Hein Schreuder, | Werknemers en Soc. Jaarverslag.  |
| 1979-2 | H. Schreuder,   | De maatschappelijke verantwoordelijkheid van Ondernemingen.  |
| 1979-3 | P. Nijkamp en P. Rietveld,                                      | Multilevel Multi-objective Models in a Multiregional System.   |
| 1979-4 | J. Arntzen, G. Bornmalm-Jardelöw and P. Nijkamp,                | Duality, Segmentation and Dynamics on a Regional Labour Market.  |
| 1979-5 | P. Nijkamp, A. Soffer   | Soft Multicriteria Decision Models for Urban Renewal Plans.  |
| 1979-6 | drs. A.J. Mathot  | A Model of choosing a car with or without a credit.  |
| 1979-7 | H. Blommestein, P. Nijkamp en W. van Veenendaal,                | Shopping Perceptions and preferences: A multi-dimensional Attractiveness Analysis of Consumer and Entrepreneurial Attitudes. |
| 1979-8 | H.J. Blommestein and F.C. Palm                                  | The Aggregate Demand for Money in the Netherlands - a new look at a study of the Bank of the Netherlands.                    |
- 
- |         |                                       |  |
|---------|---------------------------------------|--|
| 1980-1  | P. Nijkamp and H. Voogd<br>Jan.       | New Multicriteria Methods for Physical Planning by Means of Multidimensional Scaling Techniques.                           |
| 1980-2  | Hidde P. Smit                         | Medium- and Long-Term Models for the Escap-Region<br>- A review of existing models and a proposal for a new model system - |
| 1980-3  | P.v. Dijck en H. Verbruggen<br>april  | Productive Employment in Developing Countries' Exporting Industries  |
| 1980-4  | P. Nijkamp en L. Hordijk              | Integrated Approaches to Regional Development Models;<br>A survey of some Western European Models                          |
| 1980-5  | P. Nijkamp                            | Soft Econometric Models. An analysis of Regional Income Determinants   |
| 1980-6  | P. Nijkamp en F. van Dijk             | Analysis of Conflicts in Dynamical Environmental Systems via Catastrophe Theory.   |
| 1980-7  | E. Vogelvang<br>juni                  | A short term econometric model for the consumer demand of roasted coffee in the Netherlands                                |
| 1980-8  | N. van Hulst                          | De effectiviteit van de geleide loonpolitiek in Nederland.   |
| 1980-9  | P. Nijkamp<br>Oct. 1980               | A survey of Dutch integrated Energy-Environmental-Economic Policy Models   |
| 1980-10 | P. Nijkamp<br>Oct. 1980               | Perspectives for Urban analyses and policies   |
| 1980-11 | P. Nijkamp<br>Oct. 1980               | New developments in multidimensional geographical data and policy analysis   |
| 1980-12 | F.C. Palm, E. Vogelvang en D.A. Kodde | Efficient Estimation of the Geometric Distributed Lag Model; Some Monte Carlo Results on Small Sample Properties           |